Differentially Private High Dimensional Sparse Covariance Matrix Estimation[☆,☆☆]

Di Wang[a,b,*], Jinhui Xu[b]

[a]*Division of Computer, Electrical and Mathematical Sciences and Engineering*
King Abdullah University of Science and Technology
Thuwal 23955, Saudi Arabia
[b]*Department of Computer Science and Engineering*
State University of New York at Buffalo
338 Davis Hall, Buffalo, 14260

Abstract

In this paper, we study the problem of estimating the covariance matrix under differential privacy, where the underlying covariance matrix is assumed to be sparse and of high dimensions. We propose a new method, called DP-Thresholding, to achieve a non-trivial ℓ_2-norm based error bound whose dependence on the dimension drops to logarithmic instead of polynomial, it is significantly better than the existing ones, which add noise directly to the empirical covariance matrix. We also extend the ℓ_2-norm based error bound to a general ℓ_w-norm based one for any $1 \leq w \leq \infty$, and show that they share the same upper bound asymptotically. Our approach can be easily extended to local differential privacy. Experiments on the synthetic datasets show results that are consistent with theoretical claims.

Keywords: Differential privacy, sparse covariance estimation, high dimensional statistics

1. Introduction

In recent year, Machine Learning and Statistical Estimation have had profound impact on many applied domains such as social sciences, genomics, and medicine. A frequently encountered challenge in their applications is how to deal with the high dimensionality of the datasets, especially for those in genomics, educational and psychological research. A commonly adopted strategy for dealing with such an issue is to assume that the underlying structures of parameters are sparse.

[☆]A preliminary version appeared in Proceedings of The 53rd Annual Conference on Information Sciences and Systems (CISS 2019).

[☆☆]This research was supported in part by the National Science Foundation (NSF) through grants CCF-1422324 and CCF-1716400.

[*]Corresponding author

Email addresses: dwang45@buffalo.edu (Di Wang), jinhui@buffalo.edu (Jinhui Xu)

Another often encountered challenge is how to handle sensitive data, such as those in social science, biomedicine, and genomics. A promising approach is to use differentially private mechanisms for the statistical inference and learning tasks. Differential Privacy (DP) [1] is a widely-accepted criterion that provides provable protection against identification and is resilient to arbitrary auxiliary information that might be available to attackers. Since its introduction over a decade ago, a rich line of works are now available, which have made differential privacy a compelling privacy enhancing technology for many organizations, such as Uber [2], Google [3], Apple [4].

Estimating or studying the high dimensional datasets while keeping them (locally) differentially private could be quite challenging for many problems, such as sparse linear regression [5], sparse mean estimation [6], and selection problem [7]. However, there are also evidences showing that the loss of some problems under the privacy constraints can be quite small compared with their non-private counterparts. Examples of such nature include Empirical Risk Minimization under sparsity constraints [8, 9], high dimensional sparse PCA [10, 11, 12], sparse inverse covariance estimation [13], and high-dimensional distributions estimation [14]. Thus, it is desirable to determine which high dimensional problem can be learned or estimated efficiently in a private manner.

In this paper, we aim to give an answer to this question for a simple but fundamental problem in machine learning and statistics, namely estimating the underlying sparse covariance matrix of a bounded sub-Gaussian distribution. For this problem, we propose a simple but nontrivial (ϵ, δ)-DP method, DP-Thresholding, and show that the squared ℓ_w-norm error for any $1 \leq w \leq \infty$ is bounded by $O(\frac{s^2 \log p}{n} + \frac{s^2 \log p \log \frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log^2 \frac{1}{\delta}}{n^2\epsilon^4})$, where n is the sample size, p is the dimension of the underlying space and s is the sparsity of each row in the underlying covariance matrix. Moreover, our method can be easily extended to the local differential privacy model with an upper bound of $O(\frac{s^2 \log p \log \frac{1}{\delta}}{n\epsilon^2})$. Experiments on synthetic datasets confirm the theoretical claims. To our best knowledge, this is the first paper studying the problem of estimating a high dimensional sparse covariance matrix under (local) differential privacy.

2. Related Work

Recently, there have beeen several papers studying private distribution estimation, such as [14, 15, 16, 17, 18]. For distribution estimation under the central differential privacy model, [16] considers the 1-dimensional private mean estimation of a Gaussian distribution with (un)known variance. The work that is probably most closely related to ours is [14], which studies the problem of privately learning multivariate Gaussian and product distributions. The following are the main differences with ours. Firstly, our goal is to estimate the covariance of a sub-Gaussian distribution. Even though the class of distributions considered in our paper is larger than the one in [14], it has an additional assumption which requires the ℓ_2 norm of a sample of the distribution to be bounded by 1. This means that it does not include the general Gaussian distribution. Secondly, although [14] also considers the high dimensional case, it does not assume the sparsity of the underlying covariance matrix. Thus, its error bound depends on the dimensionality p polynomially, which is large in the high dimensional case ($p \gg n$),

while the dependence in our paper is only logarithmic (*i.e.,* $\log p$). Thirdly, the error in [14] is measured by the total variation distance, while it is by ℓ_w-norm in our paper. Thus, the two results are not comparable. Fourthly, it seems difficult to extend the methods of [14] to the local model. Recently, [18] also studies the covariance matrix estimation via iterative eigenvector sampling. However, their method is just for the low dimensional case and the error is measured with respect to the Frobenious norm.

Distribution estimation under local differential privacy has been studied in [17, 15]. However, both of them study only the 1-dimensional Gaussian distribution. Thus, it is quite different from the class of distributions in our paper.

In this paper, we mainly use Gaussian mechanism on the covariance matrix, which has been studied in [19, 10, 13]. However, as it will be shown later, simply outputting the perturbed covariance can incur big error and thus is insufficient for our problem. Compared to these previous work, the problem in this paper is clearly more complicated since here we assume it is in the high dimensional space where $p \gg n$.

3. Preliminaries

3.1. Differential Privacy

Differential privacy [1] is by now a de facto standard for statistical data privacy which constitutes a strong standard for privacy guarantees for algorithms on aggregate databases. DP requires that there is no significant change in the outcome distribution under a single entry change to the dataset. We say that two datasets D, D' are neighbors if they differ by only one entry, denoted as $D \sim D'$.

Definition 1 (Differential Privacy [1])**.** *A randomized algorithm $\mathcal{A}$ is (ϵ, δ)-differentially private (DP) if for all neighboring datasets D, D' and for all measurebale events S in the output space of $\mathcal{A}$, the following holds*

$$\mathbb{P}(\mathcal{A}(D) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{A}(D') \in S) + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is ϵ-differentially private.

We will use the Gaussian Mechanism [20] to guarantee (ϵ, δ)-DP.

Definition 2 (Gaussian Mechanism [20])**.** *Given any function $q : \mathcal{X}^n \to \mathbb{R}^p$, the Gaussian Mechanism is defined as:*

$$\mathcal{M}_G(D, q, \epsilon) = q(D) + Y,$$

where Y is drawn from Gaussian Distribution $\mathcal{N}(0, \sigma^2 I_p)$ with $\sigma \geq \frac{\sqrt{2\log(1.25/\delta)}\Delta_2(q)}{\epsilon}$. Here $\Delta_2(q)$ is the ℓ_2-sensitivity of the function q, i.e.

$$\Delta_2(q) = \sup_{D \sim D'} ||q(D) - q(D')||_2.$$

The Gaussian Mechanism preservers (ϵ, δ)-differential privacy.

3.2. Private Sparse Covariance Estimation

Let $x_1, x_2, \cdots, x_n$ be n random samples from a p-variate distribution with covariance matrix $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$, where the dimensionality p is assumed to be high, *i.e.,* $p \gg n \geq$ Poly(log p).

We define the parameter space of s-sparse covariance matrices as the following:

$$\mathcal{G}_0(s) = \{\Sigma = (\sigma_{ij})_{1\le i,j\le p} : \sigma_{-j,j} \text{ is } s\text{-sparse } \forall j \in [p]\}, \tag{1}$$

where $\sigma_{-j,j}$ denotes the j-th column of Σ with the entry σ_{jj} removed. That is, a matrix in $\mathcal{G}_0(s)$ has at most s non-zero off-diagonal elements in each column.

We assume that each x_i is sampled from a 0-mean and sub-Gaussian distribution with parameter σ^2, that is,

$$\mathbb{E}[x_i] = 0, \mathbb{P}\{|v^T x_i| > t\} \le e^{-\frac{t^2}{2\sigma^2}}, \forall t > 0 \text{ and } \|v\|_2 = 1. \tag{2}$$

This means that all the one-dimensional marginals of x_i have sub-Gaussian tails. We also assume that with probability 1, $\|x_i\|_2 \le 1$. We note that such assumptions are quite common in the differential privacy literature, such as [10].

Let $\mathcal{P}_p(\sigma^2, s)$ denote the set of distributions of x_i satisfying all the above conditions (*i.e.,* (2) and $\|x_i\|_2 \le 1$) and with the covariance matrix $\Sigma \in \mathcal{G}_0(s)$. The goal of private covariance estimation is to obtain an estimator Σ^{priv} of the underlying covariance matrix Σ based on $\{x_1, \cdots, x_n\} \sim P \in \mathcal{P}_p(\sigma^2, s)$ while preserving its privacy. In this paper, we will focus on (ϵ, δ)-differential privacy. We use the ℓ_2 norm to measure the difference between Σ^{priv} and Σ, *i.e.,* $\|\Sigma^{\text{priv}} - \Sigma\|_2$.

Lemma 1 ([21])**.** *Let $\{x_1, \cdots, x_n\}$ be n random variables sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Then*

$$\mathbb{E}\max_{1\le i\le n} |x_i| \le \sigma\sqrt{2\log 2n}, \tag{3}$$

$$\mathbb{P}\{\max_{1\le i\le n} |x_i| \ge t\} \le 2ne^{-\frac{t^2}{2\sigma^2}}. \tag{4}$$

Particularly, if $n = 1$, we have $\mathbb{P}\{|x_i| \ge t\} \le 2e^{-\frac{t^2}{2\sigma^2}}$.

Lemma 2 ([22])**.** *If $\{x_1, x_2, \cdots, x_n\}$ are sampled from a sub-Gaussian distribution in (2) and $\Sigma^* = (\sigma^*)_{1\le i,j\le p} = \frac{1}{n}\sum_{i=1}^n x_i x_i^T$ is the empirical covariance matrix, then there exist constants C_1 and $\gamma > 0$ such that $\forall i, j \in [p]$*

$$\mathbb{P}(|\sigma^*_{ij} - \sigma_{ij}| > t) \le C_1 e^{-nt^2\frac{8}{\gamma^2}} \tag{5}$$

for all $|t| \le \xi$, where C_1, ξ and γ are constants and depend only on σ^2. Specifically,

$$\mathbb{P}\{|\sigma^*_{ij} - \sigma_{ij}| > \gamma\sqrt{\frac{\log p}{n}}\} \le C_1 p^{-8}. \tag{6}$$

Notations:. All the constants and big-O notations throughout the paper omit the factors that are related to polynomial of σ^2, which is the sub-Gaussian parameter. Many previous papers assume the sub-Gaussian parameter as a constant, such as [22, 23].

4. Method

4.1. A First Approach

A direct way to obtain a private estimator is to perturb the empirical covariance matrix by symmetric Gaussian matrices, which has been used in previous work on private PCA, such as [19, 10]. However, as we can see bellow, this method will introduce big error.

By [19], for any given $0 < \epsilon, \delta \leq 1$ and $\{x_1, x_2, \cdots, x_n\} \sim P \in \mathcal{P}_p(\sigma^2, s)$, the following perturbation procedure is (ϵ, δ)-differentially private:

$$\tilde{\Sigma} = \Sigma^* + N = (\tilde{\sigma}_{ij})_{1 \leq i,j \leq p} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T + N, \tag{7}$$

where N is a symmetric matrix with its upper triangle (including the diagonal) being i.i.d. samples from $\mathcal{N}(0, \sigma_1^2)$; here $\sigma_1^2 = \frac{2\log(1.25/\delta)}{n^2\epsilon^2}$, and each lower triangle entry being copied from its upper triangle counterpart. By the Corollary 2.3.6 of [24], we know that $\|N\|_2 \leq O(\sqrt{p}\sigma_1) = O(\frac{\sqrt{p}\sqrt{\log\frac{1}{\delta}}}{n\epsilon})$ with high probability. We can easily get that, with high probability (*i.e.,* with probability at least $1 - \frac{1}{p^c}$ for some $c > 0$)

$$\|\tilde{\Sigma} - \Sigma\|_2 \leq \|\Sigma^* - \Sigma\|_2 + \|N\|_2 \leq O(\frac{\sqrt{p\log\frac{1}{\delta}}}{n\epsilon}), \tag{8}$$

where the second inequality is due to a Theorem in Chapter 1.6.3 of [25]. However, we can see that the upper bound of the error in (8) is quite large in the high dimensional case.

Another issue of the private estimator in (7) is that it is not clear whether it is positive-semidefinite, a property that is normally expected from an estimator.

4.2. Post-processing via Thresholding

We note that one of the reasons that the private estimator $\tilde{\Sigma}$ in (7) fails is due to the fact that some entries are quite large which make $\|\tilde{\Sigma}_{ij} - \Sigma_{ij}\|_2$ large for some i, j. More precisely, by (4) and (5) we can get the following, with probability at least $1 - Cp^{-6}$, for all $1 \leq i, j \leq p$,

$$|\tilde{\sigma}_{ij} - \sigma_{ij}| \leq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log\frac{1.25}{\delta}}\sqrt{\log p}}{n\epsilon} = O(\gamma\sqrt{\frac{\log p}{n\epsilon^2}}). \tag{9}$$

Thus, to reduce the error, a natural approach is the following. For those σ_{ij} with larger values, we keep the corresponding $\tilde{\sigma}_{ij}$ in order to make their difference less than some

threshold. For those σ_{ij} with smaller values compared with (9), since the corresponding $\tilde{\sigma}_{ij}$ may still be large, if we threshold $\tilde{\sigma}_{ij}$ to 0, we can lower the error on $\tilde{\sigma}_{ij} - \sigma_{ij}$.

Following the above thinking and the thresholding methods in [22] and [26], we propose the following DP-Thresholding method, which post-processes the perturbed covariance matrix in (7) with the threshold $\gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}$. After thresholding, we further threshold the eigenvalues of $\hat{\Sigma}$ in order to make it positive semi-definite. See Algorithm 1 for detail.

Algorithm 1 DP-Thresholding

Input: $\{x_1, x_2, \cdots, x_n\} \sim P \in \mathcal{P}_p(\sigma^2, s)$, and $\epsilon, \delta \in (0, 1)$

1: Compute

$$\tilde{\Sigma} = (\tilde{\sigma}_{ij})_{1\le i,j\le p} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T + N,$$

where N is a symmetric matrix with its upper triangle (including the diagonal) being i.i.d samples from $\mathcal{N}(0, \sigma_1^2)$; here $\sigma_1^2 = \frac{2\log(1.25/\delta)}{n^2\epsilon^2}$, and each lower triangle entry being copied from its upper triangle counterpart.

2: Define the thresholding estimator $\hat{\Sigma} = (\hat{\sigma}_{ij})_{1\le i,j\le n}$ as

$$\hat{\sigma}_{ij} = \tilde{\sigma}_{ij} \cdot I[|\tilde{\sigma}_{ij}| > \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}]. \tag{10}$$

3: Let the eigen-decomposition of $\hat{\Sigma}$ be $\hat{\Sigma} = \sum_{i=1}^{p} \lambda_i v_i v_i^T$. Let $\lambda^+ = \max\{\lambda_i, 0\}$ be the positive part of λ_i, then define $\Sigma^+ = \sum_{i=1}^{p} \lambda^+ v_i v_i^T$.

4: **return** Σ^+.

Theorem 1. *For any $0 < \epsilon, \delta \le 1$, Algorithm 1 is (ϵ, δ)-differentially private.*

Proof. By Section 3 in [19], we know that Step 1 keeps the matrix (ϵ, δ)-differentially private. Thus, Algorithm 1 is (ϵ, δ)-differentially private due to the post-processing property of differential privacy [1]. □

For the matrix $\hat{\Sigma}$ in (10) after the first step of thresholding, we have the following key lemma.

Lemma 3. *For every fixed $1 \le i, j \le p$, there exists a constant $C > 0$ such that with probability at least $1 - Cp^{-\frac{9}{2}}$, the following holds:*

$$|\hat{\sigma}_{ij} - \sigma_{ij}| \le 4\min\{|\sigma_{ij}|, \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\}. \tag{11}$$

Proof of Lemma 3. Let $\Sigma^* = (\sigma^*_{ij})_{1\le i,j\le p}$ and $N = (n_{ij})_{1\le i,j\le p}$. Define the event $A_{ij} = \{|\tilde{\sigma}_{ij}| > \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\}$. We have:

$$|\hat{\sigma}_{ij} - \sigma_{ij}| = |\sigma_{ij}| \cdot I(A_{ij}^c) + |\tilde{\sigma}_{ij} - \sigma_{ij}| \cdot I(A_{ij}). \tag{12}$$

By the triangle inequality, it is easy to see that

$$A_{ij} = \Big\{|\tilde{\sigma}_{ij} - \sigma_{ij} + \sigma_{ij}| > \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\Big\}$$
$$\subset \Big\{|\tilde{\sigma}_{ij} - \sigma_{ij}| > \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} - |\sigma_{ij}|\Big\}$$

and

$$A_{ij}^c = \Big\{|\tilde{\sigma}_{ij} - \sigma_{ij} + \sigma_{ij}| \le \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\Big\}$$
$$\subset \Big\{|\tilde{\sigma}_{ij} - \sigma_{ij}| > |\sigma_{ij}| - (\gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})\Big\}.$$

Depending on the value of σ_{ij}, we have the following three cases.

Case 1. $|\sigma_{ij}| \le \frac{\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}$. For this case, we have

$$\mathbb{P}(A_{ij}) \le \mathbb{P}(|\tilde{\sigma}_{ij} - \sigma_{ij}| > \frac{3\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}) \le C_1 p^{-\frac{9}{2}} + 2p^{-\frac{9}{2}}. \tag{13}$$

This is due to the following:

$$\mathbb{P}\big(|\tilde{\sigma}_{ij} - \sigma_{ij}| > \frac{3\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big) \tag{14}$$
$$\le \mathbb{P}\big(|\sigma_{ij}^* - \sigma_{ij}| > \frac{3\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} - |n_{ij}|\big) \tag{15}$$
$$= \mathbb{P}\big(B_{ij} \bigcap \big\{\frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} - |n_{ij}| > 0\big\}\big) \tag{16}$$
$$+ \mathbb{P}\big(B_{ij} \bigcap \big\{\frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} - |n_{ij}| \le 0\big\}\big) \tag{17}$$
$$\le \mathbb{P}(|\sigma_{ij}^* - \sigma_{ij}| > \frac{3\gamma}{4}\sqrt{\frac{\log p}{n}}) + \mathbb{P}(\frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} \le |n_{ij}|) \tag{18}$$
$$\le C_1 p^{-\frac{9}{2}} + 2p^{-\frac{9}{2}}, \tag{19}$$

where event B_{ij} denotes $B_{ij} = \{|\sigma_{ij}^* - \sigma_{ij}| > \frac{3\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{3\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}) - |n_{ij}|\}$, and the last inequality is due to (4) and (5).

Thus by (12), with probability at least $1 - C_1 p^{-\frac{9}{2}} - 2p^{-\frac{9}{2}}$, we have

$$|\hat{\sigma}_{ij} - \sigma_{ij}| = |\sigma_{ij}|,$$

which satisfies (11).

Case 2. $|\sigma_{ij}| \geq 2\gamma\sqrt{\frac{\log p}{n}} + \frac{8\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}$. For this case, we have

$$\mathbb{P}(A_{ij}^c) \leq \mathbb{P}(|\tilde{\sigma}_{ij} - \sigma_{ij}| \geq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}) \leq C_1 p^{-\frac{9}{2}} + 2p^{-8},$$

where the proof is the same as (13-17). Thus, with probability at least $1 - C_1 p^{-\frac{9}{2}} - 2p^{-8}$, we have

$$|\hat{\sigma}_{ij} - \sigma_{ij}| = |\tilde{\sigma}_{ij} - \sigma_{ij}|. \tag{20}$$

Also, by (9), (11) also holds.

Case 3. Otherwise,

$$\frac{\gamma}{4}\sqrt{\frac{\log p}{n}} + \frac{\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} \leq |\sigma_{ij}| \leq 2\gamma\sqrt{\frac{\log p}{n}} + \frac{8\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}.$$

For this case, we have

$$|\hat{\sigma}_{ij} - \sigma_{ij}| = |\sigma_{ij}| \text{ or } |\tilde{\sigma}_{ij} - \sigma_{ij}|. \tag{21}$$

When $|\sigma_{ij}| \leq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}$, we can see from (9) that with probability at least $1 - 2p^{-6} - C_1 p^{-8}$,

$$|\tilde{\sigma}_{ij} - \sigma_{ij}| \leq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon} \leq 4|\sigma_{ij}|.$$

Thus, (11) also holds.

Otherwise when $|\sigma_{ij}| \geq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}$, (11) also holds. Thus, Lemma 3 is true. □

By Lemma 3, we have the following upper bound on the ℓ_2-norm error of Σ^+.

Theorem 2. *The output Σ^+ of Algorithm 1 satisfies:*

$$\mathbb{E}\|\Sigma^+ - \Sigma\|_2^2 = O(\frac{s^2\log p}{n} + \frac{s^2\log p\log\frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log\frac{1}{\delta}}{n^2\epsilon^4}), \tag{22}$$

where the expectation is taken over the coins of the Algorithm and the randomness of $\{x_1, x_2, \cdots, x_n\}$.

Proof of Theorem 2. We first show that $\|\Sigma^+ - \Sigma\|_2 \leq 2\|\hat{\Sigma} - \Sigma\|_2$. This is due to the following

$$\begin{aligned}
&\|\Sigma^+ - \Sigma\|_2 \leq \|\Sigma^+ - \hat{\Sigma}\|_2 + \|\hat{\Sigma} - \Sigma\|_2 \leq \max_{i:\lambda_i\leq 0}|\lambda_i| + \|\hat{\Sigma} - \Sigma\|_2 \\
&\leq \max_{i:\lambda_i\leq 0}|\lambda_i - \lambda_i(\Sigma)| + \|\hat{\Sigma} - \Sigma\|_2 \leq 2\|\hat{\Sigma} - \Sigma\|_2,
\end{aligned}$$

where the third inequality is due to the fact that Σ is positive semi-definite.

This means that we only need to bound $\|\hat{\Sigma}-\Sigma\|_2$. Since $\hat{\Sigma}-\Sigma$ is symmetric, we know that $\|\hat{\Sigma}-\Sigma\|_2 \le \|\hat{\Sigma}-\Sigma\|_1$ [27]. Thus, it suffices to prove that the bound in (22) holds for $\|\hat{\Sigma}-\Sigma\|_1$.

We define event E_{ij} as

$$E_{ij} = \{|\hat{\sigma}_{ij}-\sigma_{ij}| \le 4\min\{|\sigma_{ij}|, \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\}\}. \tag{23}$$

Then, by Lemma 3, we have $\mathbb{P}(E_{ij}) \ge 1 - 2C_1 p^{-\frac{9}{2}}$.

Let $D = (d_{ij})_{1\le i,j\le p}$, where $d_{ij} = (\hat{\sigma}_{ij}-\sigma_{ij})\cdot I(E_{ij}^c)$. Then, we have

$$\begin{aligned}
&\|\hat{\Sigma}-\Sigma\|_1^2 \le \|\hat{\Sigma}-\Sigma-D+D\|_1^2\\
&\le 2\|\hat{\Sigma}-\Sigma-D\|_1^2 + 2\|D\|_1^2\\
&\le 4(\sup_j \sum_{i\neq j} |\hat{\sigma}_{ij}-\sigma_{ij}|I(E_{ij}))^2 + 2\|D\|_1^2 + O(\frac{\gamma^2\log p}{n} + \frac{\log p\log\frac{1}{\delta}}{n^2\epsilon^2}).
\end{aligned} \tag{24}$$

We first bound the first term of (24). By the definition of E_{ij} and Lemma 3, we can upper bound it by

$$\begin{aligned}
&(\sup_j \sum_{i\neq j} |\hat{\sigma}_{ij}-\sigma_{ij}|I(E_{ij}))^2\\
&\le 16(\sup_j \sum_{i\neq j} \min\{|\sigma_{ij}|, \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\})^2\\
&\le 16s^2(\gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})^2\\
&\le O(s^2\frac{\gamma^2\log p}{n} + \frac{s^2\log p\log\frac{1}{\delta}}{n^2\epsilon^2}),
\end{aligned} \tag{25}$$

where the second inequality is due to the assumption that at most s elements of $(\sigma_{ij})_{i\neq j}$ are non-zero.

For the second term in (24), we have

$$\begin{aligned}
&\mathbb{E}\|D\|_1^2 \le p\sum\nolimits_{ij}\mathbb{E}d_{ij}^2 = p\mathbb{E}\sum_{ij}[(\hat{\sigma}_{ij}-\sigma_{ij})^2 I(E_{ij}^c\bigcap\{\hat{\sigma}_{ij}=\tilde{\sigma}_{ij}\})\\
&+(\hat{\sigma}_{ij}-\sigma_{ij})^2 I(E_{ij}^c\bigcap\{\hat{\sigma}_{ij}=0\})]\\
&= p\mathbb{E}\sum_{ij}(\tilde{\sigma}_{ij}-\sigma_{ij})^2 I(E_{ij}^c) + p\sum_{ij}\mathbb{E}\sigma_{ij}^2 I(E_{ij}^c\bigcap\{\hat{\sigma}_{ij}=0\}).
\end{aligned} \tag{26}$$

For the first term in (26), we have

$$\begin{aligned}
&p\sum_{ij}\mathbb{E}\{(\tilde{\sigma}_{ij}-\sigma_{ij})^2 I(E_{ij}^c)\} \le p\sum_{ij}[\mathbb{E}(\tilde{\sigma}_{ij}-\sigma_{ij})^6]^{\frac{1}{3}}\mathbb{P}^{\frac{2}{3}}(E_{ij}^c)\\
&\le Cp\cdot p^2\frac{\log\frac{1}{\delta}}{n^2\epsilon^2}p^{-3} = O(\frac{\log\frac{1}{\delta}}{n^2\epsilon^2}),
\end{aligned} \tag{27}$$

where the first inequality is due to Hölder inequality and the second inequality is due to the fact that with some constant $C_3 > 0$,

$$\mathbb{E}(\tilde{\sigma}_{ij} - \sigma_{ij})^6 \leq C_3[\mathbb{E}(\sigma^*_{ij} - \sigma_{ij})^6 + \mathbb{E}n^6_{ij}].$$

Since n_{ij} is a Gaussian distribution, we have $\mathbb{E}n^6_{ij} \leq C_4\sigma_1^6 = O((\frac{\log\frac{1}{\delta}}{n^2\epsilon^2})^3)$ for some constant C_4 [28]. For the first term $\mathbb{E}(\sigma^*_{ij}-\sigma_{ij})^6$, since x_i is sampled from a sub-Gaussian distribution (2), by Whittle Inequality (Theorem 2 in [29] or [22]), the quadratic form σ^*_{ij} satisfies $\mathbb{E}(\sigma^*_{ij} - \sigma_{ij})^6 \leq C_5\frac{1}{n^6}$ for some positive constant $C_5 > 0$.

For the second term of (26), we have

$$\begin{aligned}
&p\sum_{ij}\mathbb{E}\sigma^2_{ij}I(E^c_{ij}\bigcap\{\hat{\sigma}_{ij}=0\})\\
&= p\sum_{ij}\mathbb{E}\sigma^2_{ij}I(|\sigma_{ij}| > 4\gamma\sqrt{\frac{\log p}{n}} + \frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})\\
&\times I(|\tilde{\sigma}_{ij}| \leq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})\\
&\leq p\sum_{ij}\mathbb{E}\sigma^2_{ij}I(|\sigma_{ij}| > 4\gamma\sqrt{\frac{\log p}{n}} + \frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})\\
&\times I(|\sigma_{ij}| - |\tilde{\sigma}_{ij} - \sigma_{ij}| \leq \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})\\
&\leq p\sum_{ij}\sigma^2_{ij}\mathbb{E}I(|\sigma_{ij}| > 4\gamma\sqrt{\frac{\log p}{n}} + \frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})I(|\tilde{\sigma}_{ij} - \sigma_{ij}| \geq \frac{3}{4}|\sigma_{ij}|)\\
&\leq p\sum_{ij}\sigma^2_{ij}\mathbb{E}I(|\sigma_{ij}| > 4\gamma\sqrt{\frac{\log p}{n}} + \frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon})I(|\sigma^*_{ij} - \sigma_{ij}| + |n_{ij}| \geq \frac{3}{4}|\sigma_{ij}|)\\
&\leq p\sum_{ij}\sigma^2_{ij}\mathbb{P}\big(\big\{|\sigma^*_{ij} - \sigma_{ij}| \geq \frac{3}{4}|\sigma_{ij}| - |n_{ij}|\big\}\bigcap\big\{|\sigma_{ij}| > 4\gamma\sqrt{\frac{\log p}{n}} + \frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big\}\big)
\end{aligned} \tag{28}$$

$$= p\sum_{ij}\sigma_{ij}^2\mathbb{P}\big(\big\{|\sigma_{ij}^*-\sigma_{ij}|\geq\frac{3}{4}|\sigma_{ij}|-|n_{ij}|\big\}\bigcap\big\{|n_{ij}|\leq\frac{1}{4}|\sigma_{ij}|\big\}\bigcap$$

$$\big\{|\sigma_{ij}|>4\gamma\sqrt{\frac{\log p}{n}}+\frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big\}\big)+p\sum_{ij}\sigma_{ij}^2\mathbb{P}\big(\big\{|\sigma_{ij}^*-\sigma_{ij}|\geq\frac{3}{4}|\sigma_{ij}|-|n_{ij}|\big\}$$

$$\bigcap\big\{|n_{ij}|\geq\frac{1}{4}|\sigma_{ij}|\big\}\bigcap\big\{|\sigma_{ij}|>4\gamma\sqrt{\frac{\log p}{n}}+\frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big\}\big) \tag{29}$$

$$\leq p\sum_{ij}\sigma_{ij}^2\mathbb{P}\big(\big\{|\sigma_{ij}^*-\sigma_{ij}|\geq\frac{1}{2}|\sigma_{ij}|\big\}\bigcap\big\{|\sigma_{ij}|>4\gamma\sqrt{\frac{\log p}{n}}+\frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big\}\big)$$

$$+p\sum_{ij}\sigma_{ij}^2\mathbb{P}\big(\big\{|n_{ij}|\geq\frac{1}{4}|\sigma_{ij}|\big\}\bigcap\big\{|\sigma_{ij}|>4\gamma\sqrt{\frac{\log p}{n}}+\frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\big\}\big). \tag{30}$$

For the second term of (30), by Lemmas 1 and 2 we have

$$p\sum_{ij}\sigma_{ij}^2\mathbb{P}(\{|n_{ij}|\geq\frac{1}{4}|\sigma_{ij}|\}\bigcap\{|\sigma_{ij}|>4\gamma\sqrt{\frac{\log p}{n}}+\frac{16\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{n\epsilon}\})$$

$$\leq p\sum_{ij}\sigma_{ij}^2\mathbb{P}(|n_{ij}|\geq\gamma\sqrt{\frac{\log p}{n}}+\frac{4\sqrt{2\log(1.25/\delta)}\log p}{n\epsilon}\})\mathbb{P}(|n_{ij}|>\frac{1}{4}\sigma_{ij})$$

$$\leq Cp\sum_{ij}\sigma_{ij}^2\exp(-\frac{(\gamma\sqrt{\frac{\log p}{n}}+4\sigma_1\sqrt{\log p})^2}{2\sigma_1^2})\exp(-\frac{\sigma_{ij}^2}{32\sigma_1^2})$$

$$\leq Cp\sum_{ij}\sigma_{ij}^2\exp(-\frac{(\gamma\sqrt{\frac{\log p}{n}}+4\sigma_1\sqrt{\log p})^2}{2\sigma_1^2})\frac{32\sigma_1^2}{\sigma_{ij}^2}$$

$$\leq C\sigma_1^2p\cdot p^2\exp(-\frac{\gamma^2\log p}{2n\sigma_1^2})p^{-8} \tag{31}$$

$$\leq C\sigma_1^2p^{-5}(\frac{2n\sigma_1^2}{\gamma^2\log p})^2=O(\frac{\log^2 1/\delta}{n^2\epsilon^4}). \tag{32}$$

For the first term of (30), by Lemma 2 we have

$$
\begin{aligned}
& p \sum_{ij} \sigma_{ij}^2 \mathbb{P}(\{|\sigma_{ij}^* - \sigma_{ij}| \geq \frac{1}{2}|\sigma_{ij}|\} \bigcap \{|\sigma_{ij}| \geq 4\gamma \sqrt{\frac{\log p}{n}}\}) \\
& \leq \frac{p}{n} \sum_{ij} n\sigma_{ij}^2 \exp(-n\frac{2\sigma_{ij}^2}{\gamma^2}) I(|\sigma_{ij}| \geq 4\gamma \sqrt{\frac{\log p}{n}}) \\
& = \frac{p}{n} \sum_{ij} [n\sigma_{ij}^2 \exp(-n\frac{\sigma_{ij}^2}{\gamma^2})] \exp(-n\frac{\sigma_{ij}^2}{\gamma^2}) I(|\sigma_{ij}| \geq 4\gamma \sqrt{\frac{\log p}{n}}) \\
& \leq \frac{p}{n} \sum_{ij} n\sigma_{ij}^2 \frac{\gamma^2}{n\sigma_{ij}^2} \exp(-16 \log p) \qquad (33) \\
& \leq C \frac{\gamma^2 p^3}{n} p^{-16} = O(\frac{1}{n}). \qquad (34)
\end{aligned}
$$

Thus in total, we have $\mathbb{E}\|D\|_1^2 = O(\frac{\log 1/\delta}{n^2\epsilon^2})$. This means that $\mathbb{E}\|\hat{\Sigma} - \Sigma\|_1^2 = O(\frac{s^2 \log p}{n} + \frac{s^2 \log p \log \frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log^2 \frac{1}{\delta}}{n^2\epsilon^4})$, which completes the proof. □

Corollary 1. *For any $1 \leq w \leq \infty$, the matrix $\hat{\Sigma}$ in (10) after the first step of thresholding satisfies*

$$
\|\hat{\Sigma} - \Sigma\|_w^2 \leq O(\frac{s^2 \log p}{n} + \frac{s^2 \log p \log \frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log^2 \frac{1}{\delta}}{n^2\epsilon^4}), \qquad (35)
$$

where the w-norm of any matrix A is defined as $\|A\|_w = \sup \frac{\|Ax\|_w}{\|x\|_w}$. Specifically, for a matrix $A = (a_{ij})_{1 \leq i,j \leq p}$, $\|A\|_1 = \sup_j \sum_i |a_{ij}|$ is the maximum absolute column sum, and $\|A\|_\infty = \sup_i \sum_j |a_{ij}|$ is the maximum absolute row sum.

Comparing the bound in the above corollary with the optimal minimax rate $\Theta(\frac{s^2 \log p}{n})$ in [22] for the non-private case, we can see that the impact of the differential privacy is an additional error of $O(\frac{s^2 \log p \log \frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log^2 \frac{1}{\delta}}{n^2\epsilon^4})$. It is an open problem to determine whether the bound in Theorem 2 is tight.

Proof of Corollary 1. By Riesz-Thorin interpolation theorem [30], we have

$$
\|A\|_w \leq \max\{\|A\|_1, \|A\|_2, \|A\|_\infty\}
$$

for any matrix A and any $1 \leq w \leq \infty$. Since $\Sigma^+ - \Sigma$ is a symmetric matrix, we have $\|\Sigma^+ - \Sigma\|_2 \leq \|\Sigma^+ - \Sigma\|_1$ and $\|\Sigma^+ - \Sigma\|_1 = \|\Sigma^+ - \Sigma\|_\infty$. Thus, by the proof of Theorem 2 we get this corollary. □

4.3. Extension to Local Differential Privacy

One advantage of our Algorithm 1 is that it can be easily extended to the local differential privacy (LDP) model.

Differential privacy in the local model. In LDP, we have a data universe $\mathcal{D}$, n players, with each holding a private data record $x_i \in \mathcal{D}$, and a server that is in charge of coordinating the protocol. An LDP protocol proceeds in T rounds. In each round, the server sends a message, which sometimes is called a query, to a subset of the players, requesting them to run a particular algorithm. Based on the queries, each player i in the subset selects an algorithm Q_i, runs it on her data, and sends the output back to the server.

Definition 3. *[31] An algorithm Q is (ϵ, δ)-locally differentially private (LDP) if for all pairs $x, x' \in \mathcal{D}$, and for all events E in the output space of Q, we have*

$$\mathbb{P}[Q(x) \in E] \leq e^{\epsilon}\mathbb{P}[Q(x') \in E] + \delta.$$

A multi-player protocol is ϵ-LDP if for all possible inputs and runs of the protocol, the transcript of player i's interaction with the server is ϵ-LDP. If $T = 1$, we say that the protocol is (ϵ, δ) non-interactive LDP.

Algorithm 2 LDP-Thresholding

Input: $\{x_1, x_2, \cdots, x_n\} \sim P \in \mathcal{P}_p(\sigma^2, s)$, and $\epsilon, \delta \in (0, 1)$

1: **for** Each $i \in [n]$ **do**

2: Denote $\tilde{x}_i\tilde{x}_i^T = x_i x_i^T + z_i$, where $z_i \in \mathbb{R}^{p\times p}$ is a symmetric matrix with its upper triangle (including the diagonal) being i.i.d samples from $\mathcal{N}(0, \sigma^2)$; here $\sigma^2 = \frac{2\log(1.25/\delta)}{\epsilon^2}$, and each lower triangle entry being copied from its upper triangle counterpart.

3: **end for**

4: Compute $\tilde{\Sigma} = (\tilde{\sigma}_{ij})_{1\leq i,j\leq p} = \frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i^T$,

5: Define the thresholding estimator $\hat{\Sigma} = (\hat{\sigma}_{ij})_{1\leq i,j\leq n}$ as

$$\hat{\sigma}_{ij} = \tilde{\sigma}_{ij} \cdot I[|\tilde{\sigma}_{ij}| > \gamma\sqrt{\frac{\log p}{n}} + \frac{4\sqrt{2\log(1.25/\delta)}\sqrt{\log p}}{\sqrt{n}\epsilon}]. \tag{36}$$

6: Let the eigen-decomposition of $\hat{\Sigma}$ be $\hat{\Sigma} = \sum_{i=1}^p \lambda_i v_i v_i^T$. Let $\lambda^+ = \max\{\lambda_i, 0\}$ be the positive part of λ_i, then define $\Sigma^+ = \sum_{i=1}^p \lambda^+ v_i v_i^T$.

7: **return** Σ^+.

Inspired by Algorithm 1, it is easy to extend our DP algorithm to the LDP model. The idea is that each X_i perturbs its covariance and aggregates the noisy version of covariance; see Algorithm 2 for detail.

The following theorem shows that the error bound of the output of Algorithm 2 is the same as the the bound in Theorem 2 asymptotically, whose proof is almost the same as in Theorem 2.

Theorem 3. *The output Σ^+ of Algorithm 2 satisfies:*

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\|_2^2 = O(\frac{s^2\log p\log\frac{1}{\delta}}{n\epsilon^2}), \tag{37}$$

where the expectation is taken over the coins of the Algorithm and the randomness of $\{x_1, x_2, \cdots, x_n\}$. *Moreover,* $\hat{\Sigma}$ *in (36) satisfies* $\|\hat{\Sigma} - \Sigma\|_w^2 = O(\frac{s \log p \log \frac{1}{\delta}}{n\epsilon^2})$.

Compared with the upper bound of $O(\frac{s^2 \log p}{n} + \frac{s^2 \log p \log \frac{1}{\delta}}{n^2\epsilon^2} + \frac{\log^2 \frac{1}{\delta}}{n^2\epsilon^4})$ in the central (ϵ, δ)-DP model, we can see that the upper bound of $O(\frac{s \log p \log \frac{1}{\delta}}{n\epsilon^2})$ in the local model is much more lower. We also note that the upper bound in the local model is tight, given by [32] recently.

5. Experiments

In this section, we evaluate the performance of Algorithm 1 and 2 in practice on synthetic datasets.

Data Generation. We first generate a symmetric sparse matrix $\tilde{U}$ with the sparsity ratio sr, that is, there are $sr \times p \times p$ non-zero entries of the matrix. Then, we let $U = \tilde{U} + \lambda I_p$ for some constant λ to make U positive semi-definite and then scale it to $U = \frac{U}{c}$ by some constant c which makes the norm of samples less than 1 (with high probability)[1]. Finally, we sample $\{x_1, \cdots, x_n\}$ from the multivariate Gaussian distribution $\mathcal{N}(0, U)$. In this paper, we set $\lambda = 50$ and $c = 200$.

Experimental Settings. To measure the performance, we compare the ℓ_1 and ℓ_2 norm of relative error, respectively. That is, $\frac{\|\Sigma^+ - U\|_2}{\|U\|_2}$ or $\frac{\|\Sigma^+ - U\|_1}{\|U\|_1}$ with the sample size n in three different settings: 1) We set $p = 100$, $\epsilon = 1$, $\delta = \frac{1}{n}$ and change the sparse ratio $sr = \{0.1, 0.2, 0.3, 0.5\}$. 2) We set $\epsilon = 1$, $\delta = \frac{1}{n}$, $sr = 0.2$, and let the dimensionality p vary in $\{50, 100, 200, 500\}$. 3) We fix $p = 200$, $\delta = \frac{1}{n}$, $sr = 0.2$ and change the privacy level as $\epsilon = \{0.1, 0.5, 1, 2\}$. We run each experiment 20 times and take the average error as the final one.

Experimental Results. Figure 1 and 2 are the results of DP-Thresholding (Algorithm 1) with ℓ_2 and ℓ_1 relative error, respectively. Figure 3 and 4 are the results of LDP-Thresholding (Algorithm 2) with ℓ_2 and ℓ_1 relative error, respectively. From the figures we can see that: 1) if the sparsity ratio is large *i.e.,* the underlying covariance matrix is more dense, the relative error will be larger, this is due to the fact that the error depends on the sparsity s, as shown in Theorem 2 and 3. 2) The dimensionality only slightly affects the relative error. That is, even if we double the value of p, the error increases only slightly. This is consistent with our theoretical analysis in Theorem 2 and 3 which says that the error of our private estimators is only logarithmically depending on p (*i.e.,* $\log p$). 3) As the privacy parameter ϵ increases (which means stronger privacy guarantees), the error becomes larger. This has also been showed in previous theorems.

In summary, all the experimental results support our theoretical analysis.

[1]Although the distribution is not bounded by 1, actually, as we see from the previous section, we can obtain the same result as long as the ℓ_2 norm of the samples is bounded by 1.

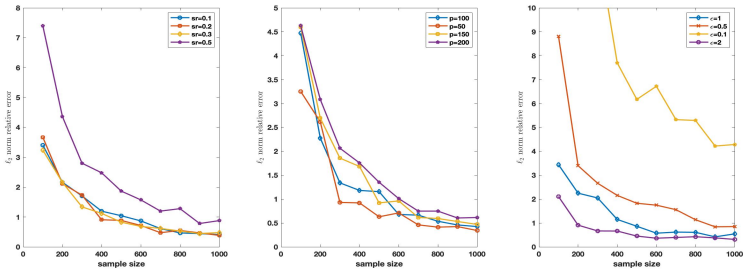

Figure 1: Experiment results of Algorithm 1 for ℓ_2-norm relative error. The left one is for different sparsity levels, the middle one is for different dimensionality p, and the right one is for different privacy level ϵ.

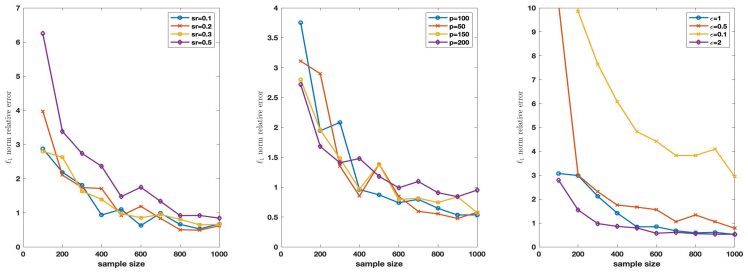

Figure 2: Experiment results of Algorithm 1 for ℓ_1-norm relative error. The left one is for different sparsity levels, the middle one is for different dimensionality p, and the right one is for different privacy level ϵ.

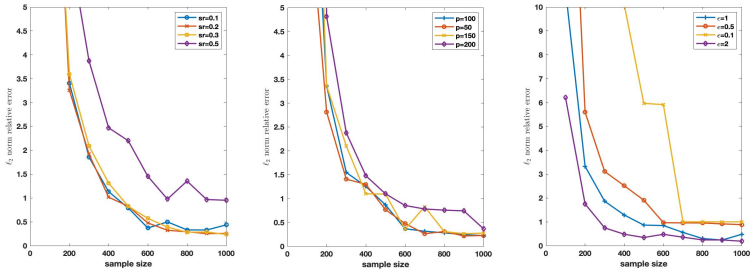

Figure 3: Experiment results of Algorithm 2 for ℓ_2-norm relative error. The left one is for different sparsity levels, the middle one is for different dimensionality p, and the right one is for different privacy level ϵ.

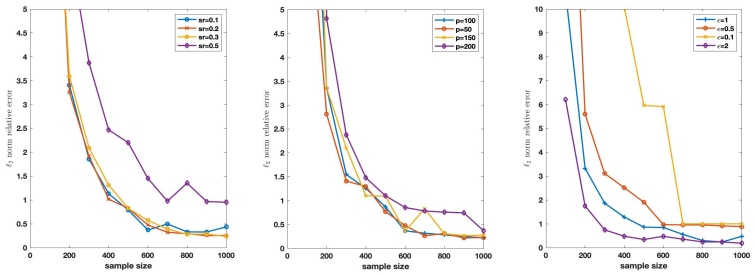

Figure 4: Experiment results of Algorithm 2 for ℓ_1-norm relative error. The left one is for different sparsity levels, the middle one is for different dimensionality p, and the right one is for different privacy level ϵ.

6. Conclusion and Discussion

In the paper, we study the problem of estimating the sparse covariance matrix of a bounded sub-Gaussian distribution in the differential privacy model and propose a method called DP-Thresholding, which achieves a non-trivial error bound and can be easily extended to the local model. Experiments on synthetic datasets yield consistent results with the theoretical analysis.

There are still some open problems for this problem. Firstly, although the thresholding method can achieve non-trivial error bound for our private estimator, in practice it is hart to find the best threshold. Thus, an open problem is how to get the best threshold. Secondly, as mentioned in the related work section, there are many recent results on private Gaussian estimation, which may make the ℓ_2 norm of the samples greater than 1. Thus, it is an interesting problem to extend our method to a general Gaussian distribution.

References

[1] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in: Theory of Cryptography Conference, Springer, 2006, pp. 265–284.

[2] J. Near, Differential privacy at scale: Uber and berkeley collaboration, in: Enigma 2018 (Enigma 2018), USENIX Association, Santa Clara, CA, 2018.

[3] Ú. Erlingsson, V. Pihur, A. Korolova, Rappor: Randomized aggregatable privacy-preserving ordinal response, in: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, ACM, 2014, pp. 1054–1067.

[4] J. Tang, A. Korolova, X. Bai, X. Wang, X. Wang, Privacy loss in apple's implementation of differential privacy on macos 10.12, CoRR abs/1709.02753. arXiv:1709.02753.

[5] D. Wang, J. Xu, On sparse linear regression in the local differential privacy model, in: International Conference on Machine Learning, 2019, pp. 6628–6637.

[6] J. C. Duchi, F. Ruan, The right complexity measure in locally private estimation: It is not the fisher information, arXiv preprint arXiv:1806.05756.

[7] J. Ullman, Tight lower bounds for locally differentially private selection, arXiv preprint arXiv:1802.02638.

[8] D. Wang, M. Ye, J. Xu, Differentially private empirical risk minimization revisited: Faster and more general, in: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, 2017, pp. 2719–2728.

[9] D. Wang, M. Gaboardi, J. Xu, Empirical risk minimization in non-interactive local differential privacy revisited, Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, 3-8 December 2018, Montreal, QC, Canada.

[10] J. Ge, Z. Wang, M. Wang, H. Liu, Minimax-optimal privacy-preserving sparse pca in distributed systems, in: International Conference on Artificial Intelligence and Statistics, 2018, pp. 1589–1598.

[11] D. Wang, J. Xu, Principal component analysis in the local differential privacy model, Theoretical Computer Science 809 (2020) 296–312.

[12] D. Wang, J. Xu, Principal component analysis in the local differential privacy model, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, AAAI Press, 2019, pp. 4795–4801.

[13] D. Wang, M. Huai, J. Xu, Differentially private sparse inverse covariance estimation, in: 2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018, Anaheim, CA, USA, November 26-29, 2018.

[14] G. Kamath, J. Li, V. Singhal, J. Ullman, Privately learning high-dimensional distributions, in: Conference on Learning Theory, 2019, pp. 1853–1902.

[15] M. Joseph, J. Kulkarni, J. Mao, S. Z. Wu, Locally private gaussian estimation, in: Advances in Neural Information Processing Systems, 2019, pp. 2984–2993.

[16] V. Karwa, S. Vadhan, Finite sample differentially private confidence intervals, in: 9th Innovations in Theoretical Computer Science Conference (ITCS 2018), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[17] M. Gaboardi, R. Rogers, O. Sheffet, Locally private mean estimation: z-test and tight confidence intervals, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 2545–2554.

[18] K. Amin, T. Dick, A. Kulesza, A. Munoz, S. Vassilvitskii, Differentially private covariance estimation, in: Advances in Neural Information Processing Systems, 2019, pp. 14213–14222.

[19] C. Dwork, K. Talwar, A. Thakurta, L. Zhang, Analyze gauss: optimal bounds for privacy-preserving principal component analysis, in: Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, 2014, pp. 11–20.

[20] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, M. Naor, Our data, ourselves: Privacy via distributed noise generation, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer, 2006, pp. 486–503.

[21] R. Vershynin, High-dimensional probability: An introduction with applications in data science, Vol. 47, Cambridge university press, 2018.

[22] T. T. Cai, H. H. Zhou, et al., Optimal rates of convergence for sparse covariance matrix estimation, The Annals of Statistics 40 (5) (2012) 2389–2420.

[23] T. T. Cai, C.-H. Zhang, H. H. Zhou, et al., Optimal rates of convergence for covariance matrix estimation, The Annals of Statistics 38 (4) (2010) 2118–2144.

[24] T. Tao, Topics in random matrix theory, Vol. 132, American Mathematical Soc., 2012.

[25] J. A. Tropp, et al., An introduction to matrix concentration inequalities, Foundations and Trends® in Machine Learning 8 (1-2) (2015) 1–230.

[26] P. J. Bickel, E. Levina, et al., Covariance regularization by thresholding, The Annals of Statistics 36 (6) (2008) 2577–2604.

[27] G. H. Golub, C. F. Van Loan, Matrix computations, Vol. 3, JHU Press, 2012.

[28] A. Papoulis, S. U. Pillai, Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, 2002.

[29] P. Whittle, Bounds for the moments of linear and quadratic forms in independent variables, Theory of Probability & Its Applications 5 (3) (1960) 302–305.

[30] N. Dunford, J. T. Schwartz, Linear operators part I: general theory, Vol. 7, Interscience publishers New York, 1958.

[31] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, A. Smith, What can we learn privately?, SIAM Journal on Computing 40 (3) (2011) 793–826.

[32] D. Wang, J. Xu, Lower bound of locally differentially private sparse covariance matrix estimation, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, AAAI Press, 2019, pp. 4788–4794.